

Middlesex University Research Repository

An open access repository of

Middlesex University research

<http://eprints.mdx.ac.uk>

Parmar, Nimai ORCID logoORCID: <https://orcid.org/0000-0001-5540-123X>, James, Nic ORCID logoORCID: <https://orcid.org/0000-0002-4181-9501>, Hearne, Gary ORCID logoORCID: <https://orcid.org/0000-0003-2146-4878> and Jones, Ben (2018) Using principal component analysis to develop performance indicators in professional rugby league. International Journal of Performance Analysis in Sport, 18 (6) . pp. 938-949. ISSN 2474-8668 [Article] (doi:10.1080/24748668.2018.1528525)

Final accepted version (with author's formatting)

This version is available at: <https://eprints.mdx.ac.uk/25291/>

Copyright:

Middlesex University Research Repository makes the University's research available electronically.

Copyright and moral rights to this work are retained by the author and/or other copyright owners unless otherwise stated. The work is supplied on the understanding that any use for commercial gain is strictly forbidden. A copy may be downloaded for personal, non-commercial, research or study without prior permission and without charge.

Works, including theses and research projects, may not be reproduced in any format or medium, or extensive quotations taken from them, or their content changed in any way, without first obtaining permission in writing from the copyright holder(s). They may not be sold or exploited commercially in any format or medium without the prior written permission of the copyright holder(s).

Full bibliographic details must be given when referring to, or quoting from full items including the author's name, the title of the work, publication details where relevant (place, publisher, date), pagination, and for theses or dissertations the awarding institution, the degree type awarded, and the date of the award.

If you believe that any material held in the repository infringes copyright law, please contact the Repository Team at Middlesex University via the following email address:

eprints@mdx.ac.uk

The item will be removed from the repository while any claim is being investigated.

See also repository copyright: re-use policy: <http://eprints.mdx.ac.uk/policies.html#copy>

Using principal component analysis to develop performance indicators in professional rugby league

Nimai Parmar^{a*}, Nic James^a, Gary Hearn^b and Ben Jones^{c,d}

^a *London Sport Institute, Faculty of Science and Technology, Middlesex University,
London, UK.*

^b *Design Engineering & Mathematics, Faculty of Science and Technology, Middlesex
University, London, UK.*

^c *Institute for Sport, Physical Activity and Leisure, Leeds Beckett University, Leeds, UK*

^d *The Rugby Football League, Leeds, UK*

**N.Parmar@mdx.ac.uk*

Using principal component analysis to develop performance indicators in professional rugby league

Previous research (Parmar et al., 2017) on performance indicators in rugby league has suggested that dimension reduction techniques should be utilised when analysing sporting datasets with a large number of variables. Forty-five rugby league team performance indicators, from all 27 rounds of the 2012, 2013 and 2014 European Super League seasons, collected by Opta, were reduced to 10 orthogonal principal components with standardised team scores produced for each component. Forced-entry logistic (match outcome) and linear (point's difference) regression models were used alongside exhaustive chi-square automatic interaction detection decision trees to determine how well each principle component predicted success. The ten principal components explained 81.8% of the variance in point's difference and classified match outcome correctly ~90% of the time. Results suggested that if a team increased 'amount of possession' and 'making quick ground' component scores, they were more likely to win ($\beta=15.6$, OR=10.1 and $\beta=7.8$, OR=13.3) respectively. Decision trees revealed that 'making quick ground' was an important predictor of match outcome followed by 'quick play' and 'amount of possession'. The use of PCA provided a useful guide on how teams can increase their chances of success by improving performances on a collection of variables, instead of analysing variables in isolation.

Keywords: rugby league; performance indicators; scoring performance; principal component analysis; regression.

Introduction

Performance indicator research is important as it allows for sport performance to be measured empirically. It has been conducted in various sports such as soccer (Castellano and Casamichana & Lago, 2012; Jones, James & Mellalieu, 2004; Lago-Peñas, Lago-Ballesteros & Rey, 2011), rugby union (Bishop & Barnes, 2013; Hughes, Hughes, Williams, James, Vučković & Locke, 2012; Villarejo, Palao, Ortega, Gomez-Ruano & Kraak, 2015) and basketball (Csataljay, O'Donoghue, Hughes & Dancs, 2009;

1 Gómez, Lorenzo, Ibañez & Sampaio, 2013; Puente, Coso & Salinero, 2015). However,
2 few papers have analysed performance indicators in rugby league (Parmar, James,
3 Hughes, Jones and Hearne, 2017; Woods, Leicht, Jones and Till, 2018; Woods, Sinclair
4 & Robertson, 2017).

5
6
7
8
9
10 Within rugby league, Woods et al. (2017) analysed three hundred and seventy-
11 six team observations taken from a publicly available statistics website, using thirteen
12 team performance indicators to assess their effect on match outcome and final league
13 position in the 2016 Australian National Rugby League (NRL) using ordinal regression
14 and conditional interference classification decision trees. Try assists, all run metres,
15 offloads, line breaks and dummy half runs were retained within the classification tree
16 detecting 66% of the losses and 91% of the wins. However, the inclusion of variables
17 such as try assists did not give meaningful information as this is simply a proxy for tries
18 scored. Future performance indicator research should exclude variables that directly
19 relate to scoring i.e. outcome variables, and rather, focus on the process variables (cf.
20 James, 2009). Furthermore, the methods indicated that Woods et al. (2017) analysed
21 team performances in isolation whereas better context could have been provided by
22 making data relative to the opposition (cf. Hughes and Bartlett, 2002). However, the use
23 of classification trees provided informative albeit simple results in regard to how
24 performances on different variables could affect success depending on the range
25 performed, which differs from regression models which only consider single unit
26 increases of the predictor variables, this becomes problematic when analysing variables
27 such as metres gained where teams sometimes perform several hundred more or less
28 than their opponents.

29
30
31
32
33
34 Within the European Super League (ESL), Parmar et al. (2017) analysed five
35 hundred and sixty-seven professional rugby league matches to determine how team
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65

1 performance indicators predicted match outcome and point's difference. The authors
2 excluded variables that related to scoring and made action variables relative by
3
4 subtracting the away team score from the home team's, therefore contextualising
5
6 performances and creating performance indicators (Hughes and Bartlett, 2002). A
7
8 combination of linear and logistic regression models alongside exhaustive chi-square
9
10 automatic interaction detection decision trees were utilised, with teams more likely to
11
12 win and gain more point's when they scored first and increased completed sets.
13
14
15 Conversely, teams decreased their chance of winning when they performed more scoots,
16
17 suggested as an inferior option compared to passing the ball.
18
19
20

21 Atkinson and Nevill (2001) recommended backwards elimination as the most
22
23 suitable regression method when analysing sport performance. This method removes
24
25 variables sequentially based on their contribution to the models' dependent variable
26
27 (e.g. match outcome). However, this comes with a reduction of predictive ability, albeit
28
29 generally small, at each step. Parmar et al. (2017) questioned the appropriateness of
30
31 stepwise methodologies when analysing sporting performance due to the removal of
32
33 variables that had previously been identified as important, and less frequently occurring
34
35 variables which could have a dramatic effect on a team's chances of winning (e.g.
36
37 yellow and red cards). They also suggested that dimension reduction techniques could
38
39 be useful for analysing large datasets that have many variables that are related to each
40
41 other, as variables that are explaining similar variance in the dataset would be grouped
42
43 together to form a component (cf. Bracewell, 2003; Parmar et al., 2017). However,
44
45 despite this, principal component analysis (PCA) has rarely been used in performance
46
47 analysis research. Rugby league performance is complex and multi-faceted and
48
49 consequently success can depend on performances on numerous variables which are
50
51 dependent on each other. Therefore, it is suggested that PCA can help produce more
52
53
54
55
56
57
58
59
60
61
62
63
64
65

1 relevant results, as it can explain that improving a set of correlated variables i.e. ball
2 carries, metres and line breaks, can lead to a higher component score, and could lead to
3 a better chance of success. Furthermore, the component scores can be calculated and run
4 in a regression model to identify how well these components can predict the variation in
5 success. This can provide coaches and analysts with more informative results to aid
6 training and tactical methods, ultimately to improve team performances.
7
8
9
10
11
12
13

14 Mackenzie and Cushion (2013) discussed the ‘theory to practice gap’,
15 suggesting that many papers lack relevance or usefulness to practitioners,
16 recommending that future performance analysis research address this issue. PCA has
17 been argued to be difficult for coaches to interpret (O’Donoghue, 2008) due to multiple
18 variables being condensed into components, however, analysing variables
19 independently of each other can also be misrepresentative as they can be related to
20 performance on other variables. For example, Woods et al. (2017) found that line breaks
21 could help determine whether a team won or lost a game. However, line breaks are
22 related to other variables such as ball carries and metres gained. Therefore, presenting
23 this variable in isolation is arguably more unrepresentative in terms of real-world
24 impact.
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40

41 Therefore, this study used PCA to reduce team performance indicators into
42 orthogonal components whilst also producing standardised scores of performance on
43 each component per game. These component scores will then be assessed using non-
44 stepwise methods due to the dataset already being reduced, hence forced-entry
45 regression analysis and decision trees will be used.
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65

Methods

Sample

Data were provided in spreadsheets (Excel v2016, Microsoft Inc., Redmond, USA) by Opta from 567 matches played in the 27 rounds of the 2012, 2013 and 2014 ESL seasons. These were extracted for analysis using Visual Basic for Applications in Microsoft Excel. Ethical approval was granted by a University Ethics Sub-Committee.

Form variables

The relative form differential between the home and away teams was assessed using five measures of form for each individual game. Five game form (point's gained in the previous 5 games) was calculated using the home team's point's minus the away team's. Similarly, current league form was calculated in the same way using total point's gained during the season.

Three further form measures used league position (end of current season league position, previous season league position and average of past three season's league position) and were calculated by subtracting the home team's league position from the away team's, hence positive values to the home team having better form and negative values when the away team had better form.

Performance indicators

Forty-five action variables (as well as form variables) were made relative by subtracting the away team's performance from the home teams, therefore creating performance indicators (PIs). Hence positive values resulted when the home team outperformed the away and negative for the opposite. PIs that related to scoring were excluded from the analysis to provide more informative results.

Statistical Analyses

PCA with orthogonal rotation (varimax method), was used to better understand the structure of the PIs and to reduce the dataset to a more manageable size to overcome multicollinearity issues in regression using IBM SPSS Statistics package (v21, IBM Corp., New York, USA). Component scores were saved using the Anderson-Rubin method. After rotation, ten components (Figure 1), which explained 73.4% of the variance were retained because of the large sample size and for having eigenvalues >1.

To enable a clear comparison of variables between winning and losing teams, draws (n=22) were excluded. The principal component scores saved from the PCA was run in both Linear (Point's difference) and Logistic (Win/Loss) forced entry regression analyses using a data splitting method (Field, 2009) on a random selection of 75% of the data. The models produced were then used to predict point's difference/match outcome using the same variables for the remaining 25% data using Minitab (v17, Minitab Inc., State College, PA). Crosstabs were performed to compare the predicted probabilities produced by the model per game to the actual match outcome. Probabilities were re-coded into winning probability (0.5-1) and losing probability (0-0.49).

Standardized residuals were analysed to ensure no bias in the regression models, if cases were within the recommended limits (Field, 2009, p.293). VIF (≤ 2.11) were not reported as there were no indications of collinearity issues (Field (2009)). Cooks distances were also analysed to ensure all values were <1 (Field, 2009) and only reported if this assumption was violated.

An exhaustive chi-square automatic interaction detection (CHAID) decision tree was grown using win/loss as the binary response variable in IBM SPSS Statistics package (v21, IBM Corp., New York, USA) using a training (75%) and test (25%) sample.

Results

PCA was conducted on forty-five PIs with the contribution of PIs to the principle component scores shown through the estimated correlations. If the PI had a positive value, it improved the component score. Conversely, when the PI had a negative value, for example missed tackles, the component score for making quick ground reduced.

FIGURE 1 NEAR HERE

Summary of Linear and Logistic forced entry regression models

The ten principal components were entered into a linear regression without stepwise methods (Table 1) which explained 81.8% of the variance in point's difference. Of the four principal components forced into the model, increasing performance on 'attempting to continue the possession' (related to both successful and unsuccessful offloads) was predicted to reduce the number of point's gained marginally. Correlation coefficients were utilised to assess models predicted point's difference and actual point's difference for forced entry linear regression models.

The same principal components were then run in a forced entry Logistic regression (Table 1). With the non-significant variables included, the predictive model suggested that having a player sent off was not likely to make a significant change in the chance of winning i.e. a 47.2% change in the probability. Similarly, the model predicted that if a team improved their 'ratio of penalties gained to conceded' principal component score by one unit, assuming all other component scores remained the same, the chances of winning would improve by 55.7% (OR=1.3). The logistic regression

1 model was able to correctly classify wins 88.3% and 90.5% of the time on the training
2 and test samples respectively.
3
4
5
6

7 **TABLE 1 NEAR HERE**
8
9

10
11
12
13 ***Residual analysis from the linear regression model***
14
15

16 One outlier was identified in the residual analysis (>3 , Table 2). The regression model
17 correctly identified the result as a home win, albeit by 11 point's as opposed to the
18 actual 44 point's. In this match, two variables (amount of possession and quick plays)
19 had values which were negative and counter to what was expected by the model.
20
21 However, the remaining four variables were consistent with the rest of the data and
22 hence the model predicted the correct result.
23
24
25
26
27
28
29
30

31
32
33 **TABLE 2 NEAR HERE**
34
35
36

37 ***Exhaustive CHAID Decision Tree (Machine Learning)***
38
39

40 A machine learning (data mining) technique was used to create a decision tree model to
41 predict winning and losing (Figure 2) from a training sample of 75%, and cross-
42 validated against a test sample of 25% of the data. The decision tree showed the most
43 important principal components being making quick ground, followed by amount of
44 possession and finally form.
45
46
47
48
49
50
51
52
53
54

55 **FIGURE 2 NEAR HERE**
56
57
58
59
60
61
62
63
64
65

1 The training sample was able to correctly classify 76.0% of games and the test-
2 sample revealed that it could classify 78.8% of games correctly (Table 8).
3
4
5

6 **Discussion**

7
8
9 The identification of variables that lead to success is an integral part of performance
10 analysis. Coaches and athletes are constantly trying to understand how to improve
11 performance, performance analysis aids this process particularly through attempting to
12 provide a greater understanding of reliable PIs and KPIs. This investigation aimed to a)
13 reduce the dataset whilst retaining as much of the variance as possible, using PCA, b)
14 assess the suitability of the principal components in predicting match outcome (logistic
15 regression and decision trees) and final point's difference (linear regression), c) provide
16 results that are relevant and transferable for practitioners.
17
18
19
20
21
22
23
24
25
26
27

28 The PCA created ten principal components, which were grouped into four main
29 categories, explaining 73.4% of the variation in the dataset. These were possession
30 (41.1%), speed of play (20.9%), form (6.0%) and infringements (5.3%), with 26.4% of
31 the variance not explained. The separation of possession and speed of play was an
32 important distinction as rugby league is a territorial game, with teams having to score by
33 moving the ball past their opponent's try line. Therefore, teams that can speed up their
34 plays are thought to gain more metres as the defending team have less time to organise
35 their defensive line adequately. The variable 'retaining possession following a kick'
36 loaded onto 'speed of play' possibly because teams that were successful on the other
37 speed variables were more successful at retaining possession following restarts,
38 logically speed would play a part in this. Success on this variable can give a significant
39 territorial advantage to a team and can easily be coached in terms of strategies to
40 maximise the potential for retaining possession. Similarly, defensive quickness can
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65

1 reduce the effectiveness of the opposition's attack ability and therefore contributes to
2 the speed of play component group.
3

4 Amount of possession loaded highly (>0.6) with metres gained and first carry
5 metres and was clearly related to gaining metres. However, the principal component
6 named "Making quick ground" loaded on variables associated with relatively dramatic,
7 sudden increases or decreases in metres gained e.g. tackle busts, support carries, missed
8 tackles and unsuccessful passes. These variables have the potential to make a significant
9 impact on the outcome of a possession as evidenced through the regression and decision
10 tree results, whilst also accounting for a large amount of variance in the dataset (20.9%)
11 and as such are key factors for coaching interventions. In addition, unsuccessful passes
12 were positively loaded onto this component. This is an unusual observation, however
13 this could simply be a proxy for a team attempting risky plays or trying to keep the ball
14 alive, which could give them a substantial advantage when performed successfully, but
15 frequently result in unsuccessful passes. The other principal components that predicted
16 significant amounts of variance were form (3a), quick plays (2c) and losing possession
17 early (1b). The "form" component was a proxy measure for individual team differences
18 which enabled the analysis to consider the differential in team qualities. However, the
19 confidence intervals are more relevant to the understanding of association. The
20 confidence intervals for form were 3.0 and 7.1 indicating that large differences in form
21 i.e. large differences in team quality, were associated with high probabilities of wins for
22 the better team whereas low positive differences associated with win probabilities were
23 akin to home advantage.
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51

52 The backwards logistic and linear regression parsimonious models both retained
53 the same five principal components, with the linear regression model also including
54 defensive quickness in its final model. This principal component consisted of one
55
56
57
58
59
60
61
62
63
64
65

1 positively loaded variable; 10m offside, where teams were caught offside at the 10m
2 mark following a tackle, more times than the opponents. An explanation could be that
3 whilst defending, the team could have a strategy of sending more players in to the tackle
4 to dominate the attacking player and prevent a quick play the ball, therefore delaying
5 the defensive retreat to the referee. On the other hand, it could be due to the team having
6 a strategy of 'line speed' where the defending team attempt to leave the line quickly to
7 prevent the opposition from gaining metres, and in the process receiving a penalty
8 against them for leaving the defensive line prior to the ball being played by the
9 opponents.
10
11
12
13
14
15
16
17
18
19
20

21 The pairwise measures of association revealed a trivial reduction in predictive
22 ability when stepwise methods were utilised. This reduction of components provided an
23 easier 'take-away message' for practitioners, however the principal components that
24 were removed could be the difference between winning and losing in closely contested
25 matches, and therefore performances on these excluded components may give teams the
26 competitive edge to win. Butterworth, O'Donoghue and Cropley (2013) mentioned the
27 potential importance of minute 'performance gains' to winning and losing on occasions
28 in sport in their review of performance profiling literature, however this approach
29 gained significant media attention after GB Cycling attributed their 7 gold medals in the
30 2012 London Olympics to their 'marginal gains' philosophy (Slater, 2012). This is
31 where they aimed to improve every component of cycling by 1%, with the collective
32 improvements resulting in better performance overall. As such it would appear that
33 stepwise procedures are not practical for analysing complex sports as variables with
34 relatively low explanatory powers are removed whereas it is reasonable to believe that
35 these could make a significant difference to a match outcome, particularly in closely
36 contested matches. Sport is a dynamic and multi-faceted process where performance
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65

1 depends on the interaction, usually reactive to the opposition, in both team and
2 individual sports.
3

4 Decision trees were utilised to identify key performance indicators, which could
5 be interpreted with ease by practitioners. Despite the transferability of results, the cross-
6 validation revealed that their predictive ability was slightly lower than the regression
7 methods. The component that explained the most amount of variance in the dataset was
8 amount of possession, which described that improving the team's ability to retain
9 possession of the ball is critical to increasing the probability of winning, which can be
10 achieved by improving the associated variables. However, the decision trees indicated
11 that making quick ground was the most important variable that could increase the home
12 team's chances of winning to 72.7%, increasing to 91.6% when also increasing the
13 amount of possession. However, large differences were evident in the confidence
14 intervals (lower 7.0 and upper 22.8) for making quick ground, this could be attributed to
15 the large variation between team qualities in the dataset. For example, top rated teams
16 would be expected to make quick ground more than lower rated teams. Furthermore,
17 each team would be expected to perform differently to each other as shown through the
18 differences in confidence intervals. Therefore, future studies could consider creating
19 nomothetic performance profiles to understand how performances on principal
20 components generally differ according to team quality and idiographic profiles for a
21 more informative understanding of individual team performances on principal
22 components.
23

24 This study found that cumulative league form loaded the highest onto the form
25 principal component, followed by final league position. This suggested that Carling,
26 Wright, Nelson and Bradley's (2014) comments, which recommended the use of current
27 form as a more appropriate and fairer method to assess team quality, were justified.
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65

Future research into other sports should consider using this measure of team quality, and assess its suitability according to the sport analysed.

The use of PCA was suggested to be a better approach to analyse performance especially when utilising regression methods which suffer from multicollinearity issues, by forming orthogonal components comprised of related variables. The suitability of this decision was evident in the residual analysis as only one outlier was identified from the large sample size. This outlier highlighted that although the team had performed on 5 components as expected, two components were unexpected (amount of possession and quick play) which led the model to predict a win by 11 point's, but which actually resulted in a winning difference of 44 point's. This case emphasised the fact that whilst sporting events can follow predictable patterns to some extent e.g. winning teams almost always gain more metres than their opponents and score more tries, large wins, as in this instance, can display unusual patterns in the data, probably due to unusual tactics which could be down to players sent off, player injuries (from either side) or poor playing conditions etc.

Future research should consider the inclusion of cumulative league form when assessing team quality, where appropriate. In addition, performances on KPIs and PIs should be assessed using performance profiling techniques to identify differences between teams. Furthermore, including the effects of independent variables when creating performance profiles is warranted, as this approach can help to provide informative results for practitioners, as this study identified large variations on some odds ratio confidence limits.

Conclusion

This study identified a method that provided a more realistic guide on how teams could

1 increase their chances of success by improving performances on a collection of
2 variables as opposed to traditional methods, which typically describe individual
3 variables. Finally, decision trees provided an insight into how machine learning can be
4 used to provide interpretable results for PCA when compared to the output from
5 regression models, despite a reduced predictive ability. Future studies could compare
6 performance on these KPIs and PIs using contextual ideographic performance profiles
7 to provide a better understanding of the variation found within and between team
8 performances on PIs and KPIs.
9
10
11
12
13
14
15
16
17
18
19
20

21 **Reference list**

- 22
23 Atkinson, G., and Nevill, A. M. (2001) 'Selected issues in the design and analysis of
24 sport performance research', *Journal of Sports Sciences*, 19(10), pp.811-827.
25
26 Bracewell, P. (2003) 'Monitoring meaningful rugby ratings', *Journal of Sports*
27 *Sciences*, 21(8), pp. 611-620.
28
29 Bishop, L. & Barnes, A. (2013) 'Performance indicators that discriminate winning and
30 losing in the knockout stages of the 2011 Rugby World Cup', *International*
31 *Journal of Performance Analysis in Sport*, 13(1), pp. 149-159.
32
33 Butterworth, A., O'Donoghue, P. and Cropley, B. (2013) 'Performance profiling in
34 sports coaching: a review', *International Journal of Performance Analysis of*
35 *Sport*, 13, pp. 572-593.
36
37 Carling, C., Wright, C., Nelson, L. and Bradley, P. (2014) 'Comment on 'Performance
38 analysis in football: A critical review and implications for future research'',
39 *Journal of Sports Sciences*, 32(1), pp. 2-7.
40
41 Castellano, J., Casamichana, D. & Lago, C. (2012) 'The Use of Match Statistics that
42 Discriminate Between Successful and Unsuccessful Soccer Teams', *Journal of*
43 *Human Kinetics*, 31, pp. 139-147.
44
45 Csataljay, G., O'Donoghue, P., Hughes, M. & Dancs, H. (2009) 'Performance indicators
46 that distinguish winning and losing teams in basketball', *International Journal of*
47 *Performance Analysis in Sport*, 9(1), pp. 60-66.
48
49 Field, A. (2009) *Discovering statistics using SPSS*. 3rd Edn. London: Sage Publications.
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65

- Gómez, M., Lorenzo, A., Ibañez, S. & Sampaio, J. (2013) 'Ball possession effectiveness in men's and women's elite basketball according to situational variables in different game periods', *Journal of Sports Sciences*, 31(14), pp. 1578-1587.
- Hughes, M., and Bartlett, R. (2002) 'The use of performance indicators in performance analysis', *Journal of Sports Sciences*, 20(10), pp. 739-754.
- Hughes, M. T., Hughes, M. D., Williams, J., James, N., Vučković, G. & Locke, D. (2012) 'Performance indicators in rugby union', *Journal of Human Sport and Exercise*, 7(7), pp. 383-401.
- James, N. (2009) 'Performance analysis of golf: Reflections on the past and a vision of the future', *International Journal of Performance Analysis in Sport*, 9(2), pp. 188-209.
- Jones, P., James, N. & Mellalieu, S. (2004) 'Possession as a performance indicator in soccer', *International Journal of Performance Analysis in Sport*, 4(1), pp. 98-102.
- Lago-Peñas, C., Lago-Ballesteros, J. & Rey, E. (2011) 'Differences in performance indicators between winning and losing teams in the UEFA Champions League', *Journal of Human Kinetics*, 27(1), pp. 135-146.
- Mackenzie, R. and Cushion, C. (2013) 'Performance analysis in football: A critical review and implications for future research', *Journal of Sports Sciences*, 31(6), pp. 639-676.
- O'Donoghue, P. (2008) 'Principal components analysis in the selection of key performance indicators in sport', *International Journal of Performance Analysis in Sport*, 8(3), pp. 145-155.
- Parmar, N., James, N., Hughes, M., Jones, H. and Hearne, G. (2017) 'Team performance indicators that predict match outcome and point's difference in professional rugby league', *International Journal of Performance Analysis in Sport*, 17(6), pp.1044-1056.
- Puente, C., Coso, J. & Salinero, J. (2015) 'Basketball performance indicators during the ACB regular season from 2003 to 2013', *International Journal of Performance Analysis in Sport*, 15(3), pp. 935-948.
- Slater, M. (2012) 'Olympics cycling: Marginal gains underpin Team GB dominance', *BBC Sport*, 8 August 2012. Available at: <http://www.bbc.co.uk/sport/olympics/19174302> (Accessed: 28 July 2017):

- 1 Villarejo, D., Palao, J., Ortega, E., Gomez-Ruano, M. & Kraak, W. (2015) 'Match-
2 related statistics discriminating between playing positions during the men's 2011
3 Rugby World Cup', *International Journal of Performance Analysis in Sport*,
4 15(1), pp. 97-111.
5
6
7 Woods, C.T., Leicht, A.S., Jones, B. and Till, K. 2018 'Game-play characteristics differ
8 between the European Super League and the National Rugby League:
9 Implications for coaching and talent recruitment'. To be published in
10 *International Journal of Sports Science & Coaching* [preprint]. Available at:
11 <http://dx.doi.org/10.1016/j.jsams.2017.04.005> (Accessed: 10 August 2018)
12
13
14 Woods, C., Sinclair, W. and Robertson, S. (2017) 'Explaining match outcome and
15 ladder position in the National Rugby League using team performance
16 indicators'. To be published in *Journal of Science and Medicine in Sport*
17 [preprint]. Available at: <http://dx.doi.org/10.1016/j.jsams.2017.04.005>
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65

Table 1. Linear and Logistic regression models

Variable	Linear Regression				Logistic Regression			
	β (SE)	β 95% CI		β (SE)	OR	OR 95% CI		Probability
		LB	UB			LB	UB	
Constant	12.0 (0.5)	11.0	13.0	1.0 (0.2)				
Amount of possession (1a)***	15.6 (0.5)	14.6	16.6	2.3 (0.3)	10.1	5.7	18.0	0.91
Making quick ground (2a)***	7.8 (0.5)	6.8	8.8	2.6 (0.3)	13.3	7.3	24.4	0.93
Form (3a)***	-2.6 (0.5)	-3.5	-1.6	1.5 (0.2)	4.7	3.0	7.3	0.82
Losing possession early (1b)***	5.1 (0.5)	4.1	6.1	-0.9 (0.2)	0.4	0.3	0.6	0.21
Quick play (2c)***	-0.3 (0.5)	-1.4	0.7	1.2 (0.2)	3.4	2.3	5.1	0.78
Attempt to continue the possession (1c)	0.7 (0.5)	-0.3	1.7	-0.3 (0.2)	0.8	0.5	1.1	0.39
Ratio of penalties gained/conceded (4a)	-0.1 (0.5)	-1.2	0.9	0.2 (0.2)	1.3	0.9	1.8	0.56
Retaining possession following a kick (2d)	1.9 (0.5)	0.9	2.9	-0.2 (0.2)	0.8	0.6	1.1	0.40
Defensive quickness (2b)	-0.3 (0.5)	-1.3	0.6	0.2 (0.2)	1.2	0.9	1.7	0.55
Player sent off (4b)	12.0 (0.5)	11.0	13.0	-0.1 (0.1)	0.9	0.7	1.3	0.47
Dataset	R ²	Pearson r with PD		Classification % Win	Classification % Loss			
Training	0.818	0.904***		88.3	83.7			
Test		0.908***		90.5	86.2			

Note: β is the unstandardized beta coefficient, SE is the standard error, OR is the odds ratio, β CI is the 95% confidence intervals for β , OR 95% CI is the 95% confidence intervals for the OR, LB is lower boundary of CI and UB is upper boundary of CI, PD is point's difference, * $p < .05$, ** $p < .01$, *** $p < .001$. Probability is probability of winning (calculation for $OR > 1 = OR/(OR+1)$); $OR < 1 = OR/2$).

Table 2. Principal component scores and model predicted values for an outlier identified in residual analysis

Case 245	Predicted	Outlier
Match outcome	Win (11 pts)	Win (44 pts)
Making quick ground (2a)	+	1.2
Amount of possession (1a)	+	-0.8
Form (3a)	+	0.4
Quick play (2c)	+	-1.8
Defensive quickness (2b)	+	1.1
Losing possession early (1b)	-	-0.3

Note: Values in red signify the divergence between the outlier values and the regression model predicted values i.e. positive (+) or negative (-)

Figure 1

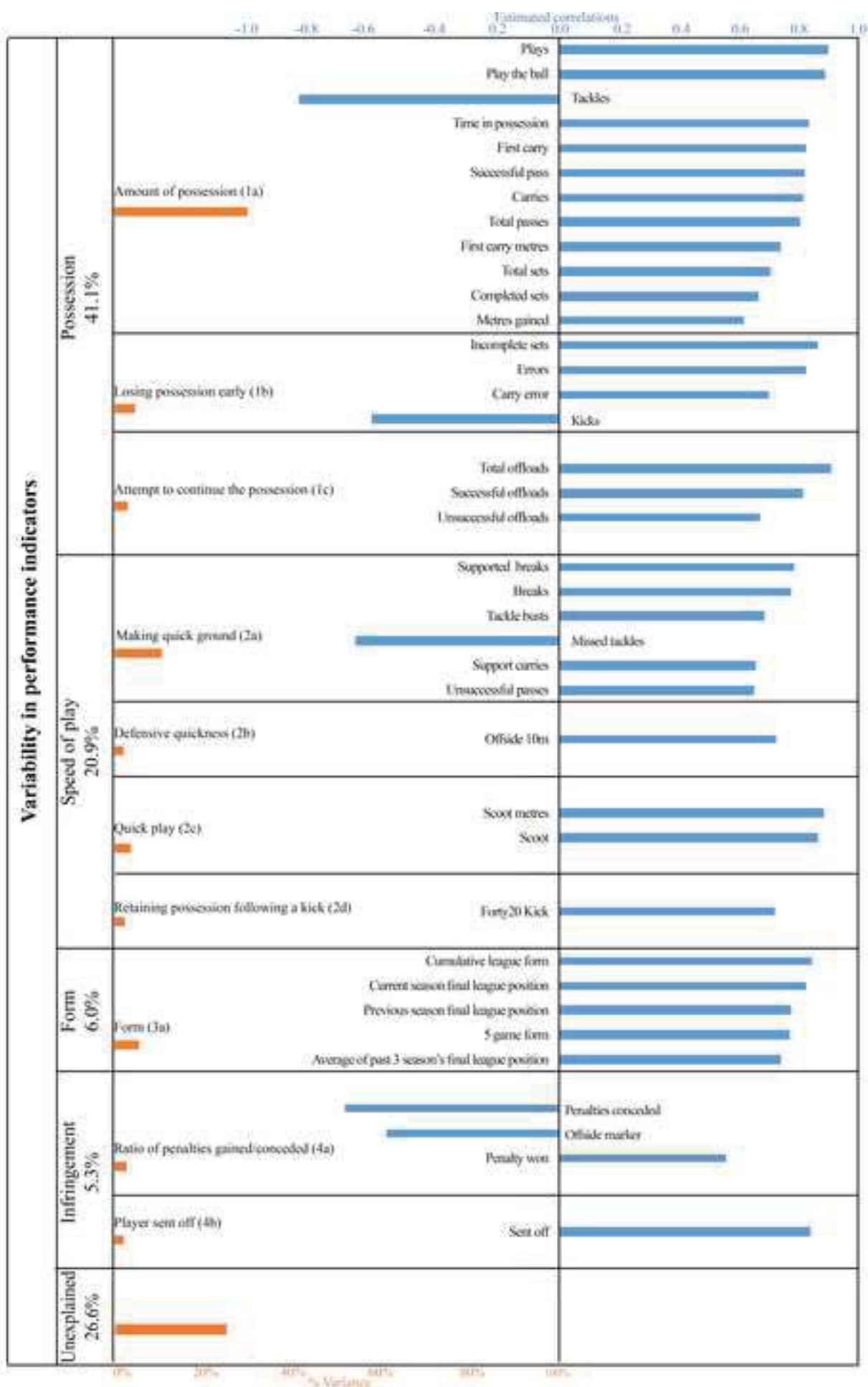


Figure 2

